Reinforcement Learning-enabled *massive* Internet of Things for 6G wireless communications

Rashid Ali[+], *Member, IEEE*, Imran Ashraf[+], Ali Kashif Bashir, *Senior Member, IEEE*, Yousaf Bin Zikria[*], *Senior Member, IEEE*

***Abstract*—Recently, extensive research efforts have been devoted to developing beyond fifth-generation (B5G), also referred to as 6^{th} generation (6G) wireless networks aimed at bringing ultra-reliable and low latency communication (URLLC) services. The 6G is expected to extend 5G capabilities to higher communication levels where numerous connected devices and sensors could operate seamlessly. One of the major research focus of 6G is to enable *massive* Internet of Things (*m*IoT) applications. Like Wi-Fi 6 (IEEE 802.11ax standard), Forthcoming wireless communication networks are likely to meet massively deployed devices and extremely new smart applications such as smart-cities for *m*IoT. However, channel scarcity is still present due to a massive number of connected devices accessing the common spectrum resources. With this expectation, next-generation Wi-Fi 6 and beyond for *m*IoT are anticipated to inherent machine intelligence capabilities to access the optimum channel resources for their performance optimization. Unfortunately, current wireless communication network standards do not support the ensuing needs of machine learning-aware (ML) frameworks in terms of resource allocation optimization. Keeping such an issue in our mind, we propose a reinforcement learning-based (RL), one of the ML techniques, a framework for wireless channel access mechanism for IEEE 802.11 standards (that is Wi-Fi) in *m*IoT. The proposed mechanism suggests exploiting a practically measured channel collision probability as a collected data-set from the wireless environment to select optimal resource allocation in *m*IoT for upcoming 6G wireless communications.**

***Index Terms*—5G, Beyond 5G, 6G, massive IoT, Resource Allocation, URLLC, Reinforcement Learning.**

I. Introduction

In recent years, significant resources are devoted by the research community towards next-generation *massive* Internet of Things (*m*IoT) wireless technologies in 5^{th} generation (5G) and Beyond 5G (B5G) networks (also referred as 6G) [1]. It is expected that the future wireless networks in *m*IoT will infer the diverse network conditions to control and optimize spectrum resources spontaneously. While cellular has its origins outdoors, we expect Wi-Fi and 6G to co-exist indoors and outdoors. IEEE working group (WG) for Wi-Fi standards (that is, IEEE 802.11 standards) has recently launched an amendment to IEEE 802.11 WLANs, named IEEE 802.11ax high-efficiency WLAN (HEW), also known as Wi-Fi 6. HEW deals with massively connected device deployment scenarios. It is anticipated that HEW infers the exciting features of both; the devices' environment and devices' interacting behavior with its environment to spontaneously manage the spectrum resource allocation. In general, a wireless device relies upon exploiting the diverse system's uncertainty in terms of transmitted data variety. Therefore, to accomplish the targeted objectives of HEW, it is imperative to examine effective and robust resource allocation schemes [2].

Today, WLANs have arrived at the time where it requires to make a change in perspective to fulfill the expanding needs of future *m*IoT applications [3]. Given the current advancement, Machine Learning (ML), especially Reinforcement Learning (RL), is expected to direct revolutionary changes, particularly concerning the spectrum resource sharing of the B5G/6G wireless communications. RL techniques are intended to engage a computational framework for learning interactively. Based on the action-state experience, future actions can be appropriately overseen without having been customized clearly. Concerning WLANs, there is an enormous measure of unexploited data created at both station (STA) and access point (AP) levels, which could be incomprehensibly essential for learning complex situations, likewise improving overall WLAN performance. For instance, the channel access experience of the STAs in a wireless network can be anticipated through RL techniques, given the information from experience. Based on these anticipations, spectrum resources can be appropriately obliged in future channel access sessions. However, RL's possible advantages for wireless networks are presently limited by the current network infrastructure, which is not yet set up to oblige RL-enabled tasks, for example, information collection, processing the information, and optimal action selection based on the processing. Instead, current wireless network frameworks are commonly implied for data transmission without considering the hidden attributes of the system.

Recently, 5G systems have initiated the moves toward ML-empowered wireless networks through network function virtualization (NFV) [4]. NFV permits fast flexibility and rapid reconfiguration in assigning spectrum resources. It is beneficial to empower verticals like self-driving cars and smart industries. Additionally, NFV is valuable to support coordination and carry ML-based operations to a large scale level, with immense data and computational complexity.

Therefore, in this paper, we acknowledge using RL-aware

R. Ali is with the School of Intelligent Mechatronics Engineering, Sejong University, Seoul, 05006, Republic of Korea, e-mails: rashidali@sejong.ac.kr.

Y. B. Zikria and I. Ashraf are with the Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, Korea, emails: yousafbinzikria@ynu.ac.kr, imranashraf@ynu.ac.kr.

A. K. Bashir is with the Department of Computing and Mathematics, Manchester Metropolitan University, Manchester, United Kingdom. He is also working with the School of Electrical Engineering and Computer Science, National University of Science and Technology(NUST), Islamabad, Pakistan, e-mail: dr.alikashif.b@ieee.org.

[+] R. Ali and I. Ashraf are equally contributed First authors.

[*]Corresponding author: yousafbinzikria@ynu.ac.kr

frameworks for next-generation WLAN networks, such as 802.11be and beyond, to add the advancement of wireless communications toward ML-based frameworks, which will be a fundamental part of the 6G wireless communications. In contrast to the mobile cellular networks like 5G, HEW networks have gotten considerably less consideration when planning ML-aware solutions and applications. The reason is that mobile phone networks fit perfectly with big data analytic because of the enormous measure of information and high computational resources available for cellular network operators. On the other hand, HEW represents a set of explicit issues due to their dense deployment scenarios, such as train station, stadium university campus, etc., and their typical distributed nature. However, despite the truth that HEW can tally with plenty of information to be utilized by ML techniques in massive deployments, we find other resource constraint situations, like residential-type deployment. In these cases, tremendous computing and processing resources for spectrum access can not be provided to the ML activity.

The RL module-based framework permits adapting to the problem instance and the set of available resources in the environment to empower the incorporation of ML-aware methodologies into WLANs' various modalities, thus giving adaptability in terms of dense deployment heterogeneity. For example, despite deep learning is a ground-breaking solution that may improve the network performance in different situations, it involves many computations, massive data storage, and ultra-reliable and low latency communication (URLLC) requirements to be satisfied in various deployments or parts of the network. Though, in an RL technique, a learner learns the actions in its surrounding environment to maximize its expected reward for the corresponding actions. The learner learns the optimal policies and actions to map current states for unknown future states in the environment. The states, action, rewards, and state-transition probabilities depict the new environment. It makes it evident for RL-aware frameworks to fit for next-generation wireless communications perfectly.

Following are the main contributions of this paper:

- This paper devises and examines the capability of RL-empowered future communications. At that point, we focus on IEEE 802.11 WLANs for efficient spectrum access.
- This paper provides an overview of the RL-aware architecture for next-generation wireless communications.
- We portray the expected advantages of RL-based methodologies empowered by the proposed framework through simulation results in a particular use case.

II. Machine Learning as an Incumbent Technology for next-generation WLANs

A brief discussion is required to elaborate on ML techniques' critical role in supporting next-generation WLANs' advancement. In this section, we specifically focus on the application of ML to next-generation 802.11 networks, that is, IEEE 802.11be and beyond.

The advancement of next-generation communication applications is characterizing the shape of future WLANs through a bunch of strict prerequisites [4]. A few models are Vehicle to Everything (V2X), Industry 4.0, and Virtual Reality/Augmented Reality (VR/AR) in 6G communications. These applications are truly challenging regarding transmission capacity (that is, a bandwidth of 10-20 Gbps), less than $5ms$ latency, 99.9% reliability, and scalability of 1,000,000 $devices/km^2$. In 5G, the advanced technologies are included, such as Enhanced Mobile Broadband (eMBB), Massive Machine to Machine Communication (mMTC), and URLLC. Similarly, 802.11 WG are also considering these technologies to design next-generation advancements, such as IEEE 802.11ax HEW and IEEE 802.11be Extreme High Throughput (EHT).

To meet the previously mentioned existing requirements, not just a technological advancement is required (e.g., utilization of higher spectrum or massive antennas technologies), yet a paradigm shift is essential when planning novel solutions for communication frameworks, operation, and management. Specifically, AI-enabled wireless communications need to be engaged with cognitive (behaviorist) and context-aware abilities, which may require a novel framework. Keeping this in mind, ML is required to be significant during the lifetime of 5G and will get inescapable as included from the earliest starting point in their origination for 6G communications.

The genuine utility of ML lies in those issues that are difficult to tackle by conventional frameworks because of their intricate underlying patterns (e.g., network density and traffic load estimation). Various ML techniques have been classified into multiple ways. However, the most widely recognized taxonomy differentiates supervised learning (SL), unsupervised learning (uSL), and RL. In SL, labeled data is used for training an agent. The uSL requires no input data labels, whereas RL uses exploration and exploitation trade-off with labeled and unlabeled input data. Fig. 1 shows a few of the algorithms and potential wireless communication applications for each kind of ML algorithm, along with examples of inputs required by these techniques. We assume the additional discussion on these ML categories and techniques is out of the scope of this paper, and we suggest readers refer to the references [5]–[7] for further details.

In addition to the specific ML-enabled solutions for wireless communications issues, few efforts have been made toward empowering ML-aware frameworks in more general terms. Specifically, several framework recommendations have been proposed so far [8]–[10]. In addition to the specific ML-enabled solutions for wireless communications issues, few efforts have been made toward empowering ML-aware frameworks in more general terms. Specifically, several framework recommendations have been proposed so far [8]–[11]. The majority of the related research works concede to the vital necessity for empowering data analytic in network deployments, possibly at the stations' (STA) and access points' (AP): data gathering, data preparation, analyzing the data, and finally, future actions selections based on the analysis. In this regard, we look deeper into RL operation and focus the actual strategies, including data gathering, analyzing, and optimal action selection.

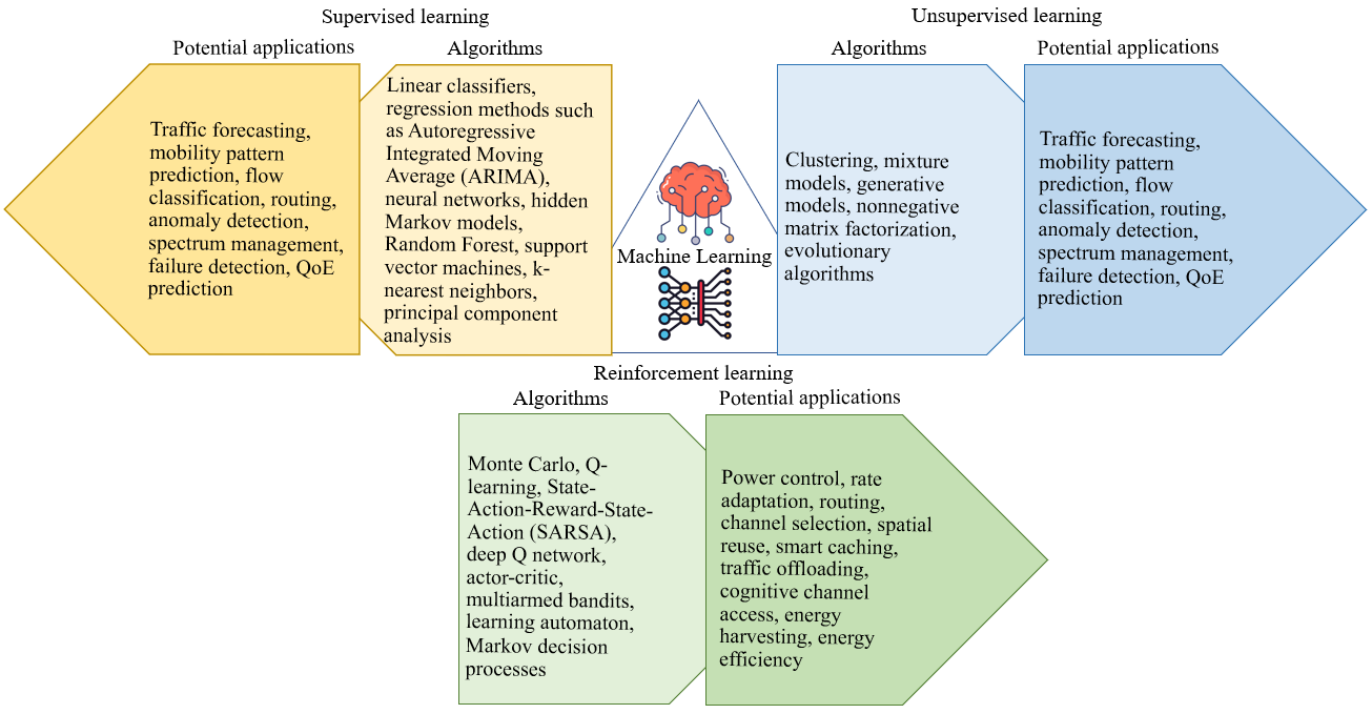

Fig. 1. Machine learning categories, algorithms, and potential communication applications.

A. ML-enabled use cases for wireless networks

It is essential to describe use cases where ML-enabled applications improve network performance. Therefore, in this section, we discuss a set of ML-enabled use cases to showcase the potential of ML in next-generation 802.11 networks.

1) Network Slicing: Network Slicing (NS) is probably the most sweltering research topic in 5G communications due to its capability to virtually isolate network resources to meet diverse application necessities. In future WLANs, NS can be realized through the optimal resource allocation of spectrum resources using orthogonal frequency-division multiple access (OFDMA). However, the diversity of applications and devices and their subsequent flexibility become the challenge for easily allocating spectrum resources. To tackle this, ML can be utilized to predict the user requirements for network performance optimization.

2) Handover and Association management: The greater part of the current user association and handover techniques in wireless networks typically depends on signal strengths. It may be challenging as load balancing can lead to serious performance degradation in densely deployed wireless networks like HEW. Thus, an ML-aware framework is conceivable to deal with context-oriented data, such as the traffic load, to help optimal action selection. Besides, user mobility and requirements prediction can be included in the framework, consequently empowering the handover and user association management with insightful data.

3) Coordinated Scheduling: Contrary to the conventional cellular communication networks, a HEW deployment can be denser, particularly in a public residential situation where anyone can set up an AP and make their wireless network. It usually prompts more complex situations where BSS collaborations prevent the current scheduling techniques from ensuring QoS. Thus, ML can be utilized to induce these interactions and bring optimal coordinated scheduling. Specifically, through ML-enabled coordinated scheduling, diverse APs can trigger uplink/downlink transmissions from/to the proper STAs, increasing the overall network throughput while lowering the channel collision among the STAs.

4) Spatial Re-use: Spatial re-use (SR) targets to improve spectrum utilization through channel sensitivity adjustment techniques. However, choosing the optimal channel sensitivity threshold limit is very difficult due to the complex spatial communications among the STAs. At this point, as a potential framework, RL techniques can be applied locally to improve spectral resource allocation in a decentralized and distributed way.

III. Reinforcement Learning-Aware Framework for IEEE 802.11 WLANs

In the RL algorithm, an agent performs actions within a state of its environment to collect a value/rewards as shown in Fig. 2. A typical RL technique has three key sub-components; strategy or policy, reward, and value function (most of the time referred to as Q-value function) [12]. The policy is a key component in the RL technique, and it characterizes a strategy for a learning agent to behave in its environment. Also, with each action, an agent earns a reward from the system. The reward value is a numerical value, and the main objective of an agent is to maximize its accumulated reward for any specific action-state pair. Similarly, a value function (or a Q-value function) represents a long-term accumulated reward for a given action. The instant reward for a specific action might be small, but it can be a higher value of a

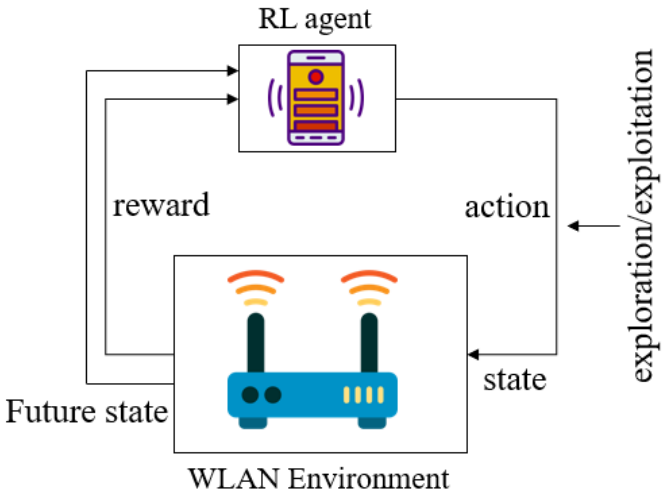

Fig. 2. Interaction of a typical agent of RL technique with its environment.

TABLE I
MAC/PHY LAYER SIMULATION PARAMETERS FOR PERFORMANCE EVALUATION.

Parameter Type	Value
Frequency	$5GHz$
Channel bandwidth	$160MHz$
Data rate (MCS11)	$1201Mbps$
Payload size	$1472bytes$
Transmission range	$10m$
CW_{min}	32
CW_{max}	1024
Simulation time	$500sec$
Propagation loss	LogDistancePropagation
Mobility	ConstantPositionMobility
Rate-adaptation	ConstantRateWifiManager
Error-rate	NistErrorRateModel

Q-value function in the long run. It indicates that a high-value action is visited several times by the agent due to the action's exploitation as an optimal action. The Q-learning is one of the algorithms of model-free RL techniques to solve behaviorist decision problems. It uses learning rate to adjust the learning capability, discount factor to give higher/lower value to the future reward, instant reward, and change in Q-value function to update current Q-value function. The maximization of the Q-value function leads to the optimal action selection in the environment. The Q-learning strategy has been used successfully in the optimization of cognitive radios and wireless channel access techniques.

A. Realization of RL-aware Framework for WLANs

To exhibit the RL-aware framework's appropriation for WLANs, let us take the example of channel observation-based spectrum resource allocation [13]. We propose a hybrid RL-aware solution where two principle RL-based processes are performed: training the model (learning from the practical channel information) and placement of the model (optimize the resource allocation based on the learned information). Fig. 3 represents the key stages of the proposed RL-aware framework for WLANs in an *m*IoT environment.

While training of the model is done at the AP with the collection of channel observation data from numerous STA, the model's placement is also done at the AP to provide an immediate response to future actions (exploitation). Notice that the framework can likewise be re-trained during the second stage based on newly explored observation data (exploration).

1) Training Phase: In our proposed framework, the STAs in a wireless environment observe the channel for channel observation-based collision probability as in [13] (as shown in red in Fig. 3). Later, the AP gathers this data of various STAs during their uplink transmission. The channel collision probability can be utilized for either training or algorithms that help the fundamental MAC layer resources allocation (MAC-RA), such as optimal contention window selection [14]. The AP's collected data is pre-processed with the goal that the RL technique can appropriately learn the channel conditions. For example, in the case of applying a Q-learning [12], the input data needs to be converted into the value-based information as rewards (that is, convert the channel observation information into a collision probability of a scalar between 0 and 1). While generating the RL framework, certain rules should be considered. For example, based on the spectrum resources, an AP may set a maximum number of connected STAs. The rules are strongly attached to the abilities of the wireless devices. Once the RL strategy at the AP generates the output (that is, the optimized MAC-RA function), it is distributed throughout the network environment to the STAs, which are then prepared to give fast optimal spectrum resource allocation to new cases.

2) Placement Phase: In the placement phase (as shown in blue in Fig.32), an AP can detect new spectrum resource requests or potential handovers based on recently collected data from STAs. The collected data is processed by the AP, similarly as in the training phase. The Q-learning technique is applied locally at the AP, which provides a reward-based output for future requests. The MAC-RA decision is conveyed to the associated STAs.

B. Potentials of RL-based Framework

To feature the capability of the RL-based framework through simulation results, we compare the throughput performance of the conventional MAC-RA (ConMAC-RA) mechanism with a channel observation-based MAC-RA (COBMAC-RA) mechanism and a novel RL-based MAC-RA (RLMAC-RA) approach [14]. We performed simulations in network simulator 3 (NS-3) [15]. Table I lists all the simulation parameters used for the performance evaluation of the RL-aware framework. Specifically, the RLMAC-RA predicts the throughput that an STA will acquire after the association with a given AP based on a channel observation-based collision probability information. Fig. 4(a) shows the network throughput for a different number of connected STAs. We see that the RLMAC-RA approach improves the average throughput performance and optimizes the over-all network performance to allow a much greater number of STAs within the environment. Similarly, in Fig. 4(b), we increase the number of connected STAs within the same environment with time. The figure shows an RLMAC-RA mechanism that learns the

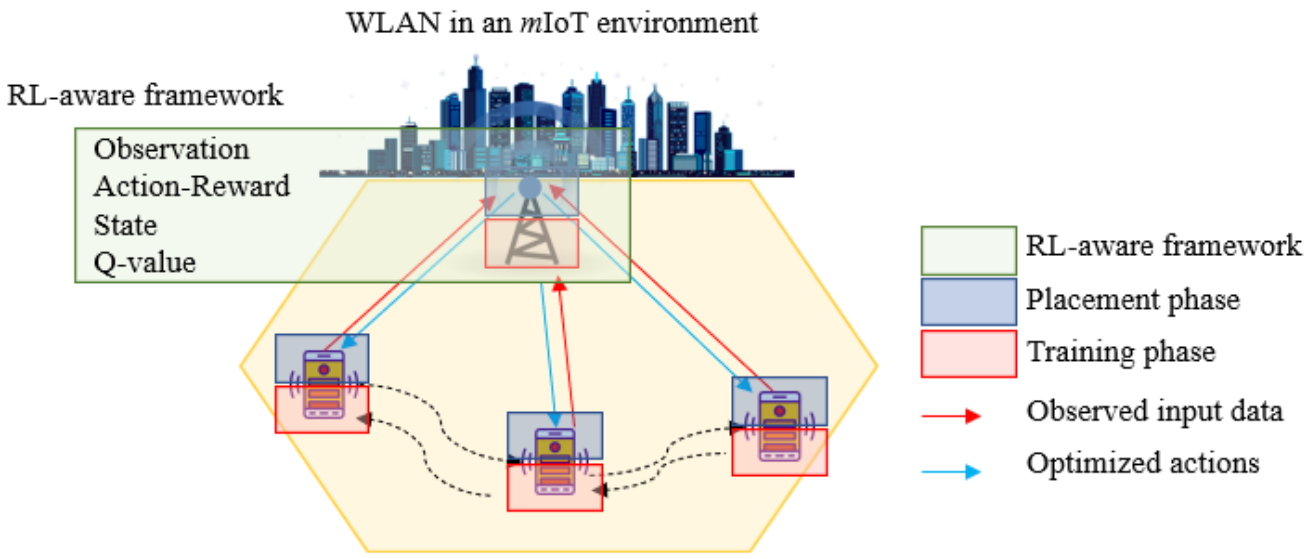

Fig. 3. Key stages of the proposed RL-aware framework for WLANs in an *m*IoT environment.

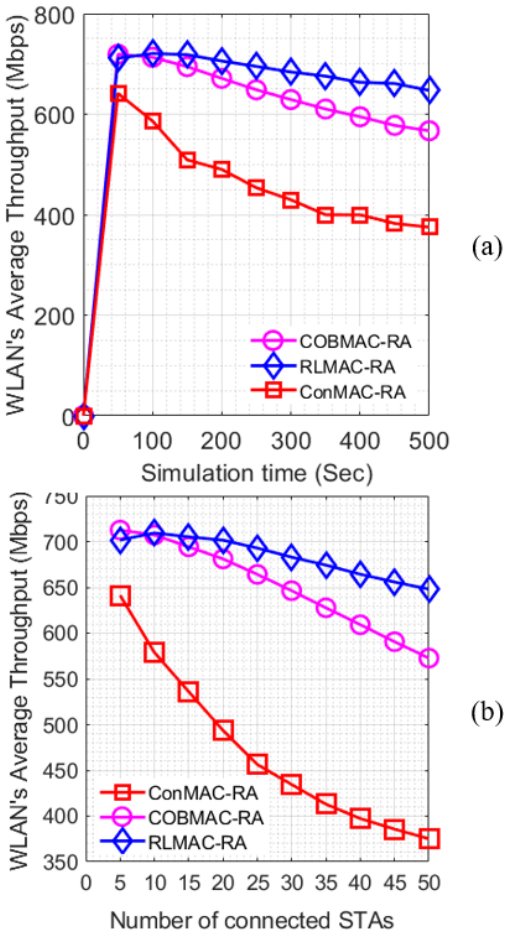

Fig. 4. Throughput comparison among ConMAC-RA, COBMAC-RA, and RLMAC-RA in
a) WLAN's average throughput comparison for a different number of connected STAs. b) a dynamic network environment with an increasing number of connected STAs after every 50 seconds.

environment and converges the system throughput to the optimal level. The RL technique can interactively learn complex and dynamic situations from dense deployments, consequently ensuring optimal throughput requirements to STAs. One of these figures' interesting observations indicates that an RL-aware framework for spectrum resource allocation may allow many connected devices within a WLAN environment. As shown in Fig. 4(a) and Fig. 4(b), the network throughput of the conventional MAC-RA (ConMAC-RA) mechanism degrades with the increase of several connected STAs, resulting in a very low or possibly zero throughput in the network due to increased collisions among the STAs. On the other hand, the RL-aware MAC-RA (RLMAC-RA) mechanism is more stable and converged even with the number of connected STAs within the network.

IV. Conclusion

Current Wireless communication networks, like IEEE 802.11 standards, are not yet ready for the pervasive adoption of ML-based frameworks. Therefore, disruptive framework level changes are required for upcoming wireless communication standards. This article presents an RL-aware framework for next-generation wireless communications o cope with such a situation in future technologies, 5G and beyond (6G) for IEEE 802.11 WLANs (such as IEEE 802.11ax). Our proposed framework provides enhanced network performance in throughput and allows a WLAN network to support many connected STAs.

Thus, we conclude that future WLANs are imagined sharing a typical flexible RL-aware architecture that permits optimized spectrum resources allocation. Nevertheless, plenty of efforts are yet required before arriving at knowledgeable wireless networks. We highlighted an RL-based framework for data handling (collection from the WLAN environment), coordi-

nation (distribution of the RL operation and dealing with the data heterogeneity), and robustness of the RL strategies (managing vulnerability and preventing the exceptional events in the environment).

REFERENCES

[1] K. Sheth, K. Patel, H. Shah, S. Tanwar, R. Gupta and N. Kumar, "A taxonomy of AI techniques for 6G communication networks," Computer Communications, Vol. 161, pp. 279-303, 2020. DOI:10.1016/j.comcom.2020.07.035.

[2] R. Ali, S. W. Kim, B. Kim and Park Y. "Design of MAC Layer Resource Allocation Schemes for IEEE 802.11ax: Future Directions," *IETE Technical Review*, 2016, **35(1)**, pp. 28-56. doi: 10.1080/02564602.2016.1242387.

[3] Afif Osseiran et al., "Scenarios for 5G mobile and wireless communications: the vision of the METIS project," IEEE Communications Magazine, vol. 52, no. 5, pp. 26–35, 2014.

[4] ITU-T Supp. Y.Supp55, "Machine learning in future networks including IMT-2020: use cases," 2019.

[5] Chunxiao Jiang et al., "Machine Learning Paradigms for Next-Generation Wireless Networks," IEEE Wireless Commun., vol. 24, no. 2, Apr. 2016, pp. 98–105.

[6] C. Zhang, P. Patras, and H. Haddadi, "Deep Learning in Mobile and Wireless Networking: A survey," IEEE Commun. Surveys & Tutorials, vol. 21, no. 3, 2019, pp. 2224–87.

[7] U. Muhammad et al., "Unsupervised Machine Learning for Networking: Techniques, Applications and Research Challenges," IEEE Access, vol. 7, 2019, pp. 65 579-65 615.

[8] Suzhi Bi et al., "Wireless Communications in the Era of Big Data," IEEE Commun. Mag., vol. 53, no. 10, Oct. 2015, pp. 190-99.

[9] I. Chih-Lin et al., "The Big-Data-Driven Intelligent Wireless Network: Architecture, Use Cases, Solutions, and Future Trends," IEEE Vehic. Tech. Mag., vol. 12, no. 4, 2017, pp. 20-29.

[10] M. Wang et al., "Machine Learning for Networking: Workflow, Advances, and Opportunities," IEEE Network, vol. 32, no. 2, Mar./Apr. 2018, pp. 92-99.

[11] M. Sohail, R. Ali, M. Kashif, S. A. Khanand, S. Mehta, Y. B. Zikria, and H. Yu, "TrustWalker: An Efficient Trust Assessment in Vehicular Internet of Things (VIoT) with Security Consideration," Sensors 2020, 20(14), 3945.

[12] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 1998.

[13] R. Ali, N. Shahin, Y. Kim, B. Kim and S. W. Kim, "Channel observation-based scaled backoff mechanism for high-efficiency WLANs," *Electronics Letters*, May 2018, **54(10)**, pp. 663-665. doi: 10.1049/el.2018.0617.

[14] R. Ali, N. Shahin, Y. B. Zikria, B. Kim and S. W. Kim, "Deep Reinforcement Learning Paradigm for Performance Optimization of Channel Observation-based MAC Protocols in Dense WLANs," IEEE Access, vol. 7(1), pp. 3500-3511, January 2019.

[15] The Network Simulator-ns-3. [Online Available at]: https://www.nsnam.org/ [Accessed on: 01-06-2020].